I can see it; why can't the computer?

Tracking humans: the need for engineering solutions

Nicola J. Ferrier
Robotics Lab
Univ. of Wisconsin-Madison

Outline of Talk

- Technology for acquiring 3D
  - Can digital video work?
- Visual Tracking (of Humans)
  - What is visual tracking?
  - Visual tracking modalities.
- 3D reconstruction
  - Correspondence
  - Changes in shape, illumination, etc.
  - Accuracy / robustness / precision

3D motion-data acquisition technologies

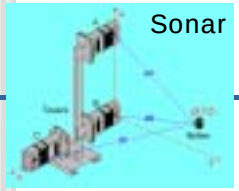
- Mechanical Armatures
- Active Marker Tracking
- Passive Marker Tracking

Mechanical Armatures

- Invasive
- Expensive
- Accurate
- Large Working Volume

Fibre Optic ShapeTape©

Sonar



Active Marker Tracking

Infrared

Electromagnetic

Passive Marker

- Infrared

- Digital Video ?

Can video provide adequate 3D information for motion capture applications?

It would be desirable to use video because it's inexpensive, portable, and non-invasive.

This session is intended to assess how realistic this is.

Must locate, track objects, then reconstruct

What is Visual Tracking?

- Determine geometric attributes of a target via measurements obtained from a sequence of images (usually temporally ordered).
- Given:
  - target in an image, I_0
  - initial configuration (or state), s_0, of the target
  - sequence of subsequent images, I_1, I_2, …
- Find:
  - a corresponding series of estimates of the target location/configuration $s_1, s_2, \ldots$

Why is it Hard?

- Variability
  - Pose
  - Illumination
  - Deformations or articulations
  - Full or partial occlusions
- Distractors

Distractors






Basic Questions

- What cues define the object of interest?
- What viewing conditions can be expected?
  - Changes in pose, illumination, etc
- What prior information can be assumed?
  - Skeletal models, dynamic models
- What level of information is needed?
  - Accuracy, frequency of measurements

Tracking modalities are :
Blobs, Templates, Regions, Contours…

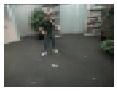
Blob Tracking

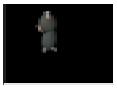
- Segment a region of interest based on intensity, color, motion, texture, color histograms and/or depth

Identify a region of interest R_0 in an initial image, I_0
with center of the region at image location I_{u1}
For every image I_1, I_2, … ,
(a) acquire R_k about I_{uk-1} in I_k
(b) Evaluate the segmentation function on R_k:
B_k = pixels in blob
(c) compute $I_{uk} = I_{uk-1}$ + centroid(B_k)

Tracking People as Blobs

- Skin finding, Eye Finding
- Pfinder
  - [Wren, Azarbayejani, Darrell, Pentland, 1997]

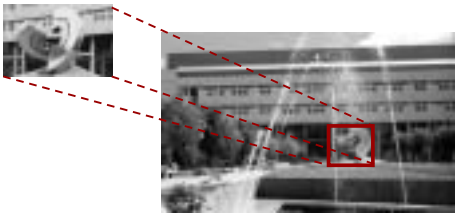




Template Tracking

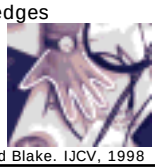
- Choose a region-of-interest, R_0 in I_0 at location I_{u0}
- For every image I_k:
  i. Use some variation on correlation/sum of squared differences to find the location, I_{uk}, of the best match to R_{k-1} in I_k.
  ii. Sample R_k from I_k about I_{uk}.

Template Matching



Contour/Snake Tracking

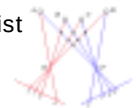
- Identify an initial contour, C_0, in an initial image, I_0.
- for each subsequent image I_k
  - traverse the current contour along a set of discrete points,
  - at each s_j, search orthogonal to the contour for an edgel (intensity edges or motion edges).
  - Estimate a new contour C_k in I_k using the information using, for example, a least squares fit.



Blake, Curwen, & Zisserman. IJCV 1993. Isard and Blake. IJCV, 1998

3D From Visual Tracking

- Track in two or more cameras
  - Correspondence must be solved
  - Cameras must be calibrated
- Use 3D model of object or kinematics (e.g. cars [Koller, Bregler]
  - Multiple solution branches exist (Gleicher & Ferrier, 2001)



3D Model Based Methods

- Statistical distributions [Sdienbladh & Black]

Summary

- Tracking from video is possible subject to satisfying the assumptions of the tracking method.
- Using the tracking date to get 3D requires explicit models of the camera/target etc.